

# Steven Palayew

(613) 864-6808 | spalayew@gmail.com  
linkedin.com/in/steven-palayew

## Education:

### Ph.D., Computer Science

January 2025 – August 2028 (Expected)

*University of Toronto*

- Supervised by Dr. Rahul G. Krishnan, Dr. Michael Wainberg, and Dr. Bo Wang.
- Other Affiliations: Vector Institute, Lunenfeld-Tanenbaum Research Institute.
- Research on applying large language models in genomics, with a focus on multimodality, reasoning, and drug discovery.

### M.Sc., Computer Science

September 2023 - January 2025

*University of Toronto*

- Supervised by Dr. Gary Bader and Dr. Bo Wang.
- Completed thesis on applying large language models (LLMs) to complement single-cell foundation models. Used interpretability methods to elucidate what knowledge of biology LLMs leverage when applied to single-cell analysis, then developed methods to leverage synergies between LLMs and single-cell foundation models, improving cell type classification and disease phenotype prediction performance.
- Collaborated with a team of two other graduate students to win both MEDIQA 2024 shared task competitions hosted by NAACL-ClinicalNLP 2024. Presented findings at conference.
- Teaching Assistant for *CSC108: Introduction to Computer Programming* (Winter 2024).

### B.A.Sc., Honours Mechatronics Engineering

September 2018 - June 2023

*University of Waterloo*

- Computing Option.
- University of Waterloo President's Scholarship of Distinction Recipient.

**Research Interests:** Large Language Models, Computational Biology, Multimodality, AI for Healthcare.

## Technical Skills:

- Proficient in Python, R, C/C++, MATLAB.
- Developed multiple projects using TensorFlow, PyTorch, Keras, OpenAI API, LangChain, Hugging Face Transformers, Scikit-Learn, SciPy, Pandas, Linux CLI, AWS, and Git.

## Professional Experience:

### Deep Learning Developer

May 2022 - August 2022

*DarwinAI (Acquired by Apple Inc.), Waterloo, Ontario*

- Brought a project recognized by the United Nations for innovatively leveraging AI in sustainable agriculture from an initial proof of concept, to a minimum viable product. This was accomplished by moving to a constrained differential evolution based approach, which improved yield optimization performance by approximately 40-50%, and ensured proposed environmental conditions were realistic. In addition, deployed the project as a web app to gather customer feedback.
- Co-first authored a paper involving the integration of a novel backbone into RetinaNet. Main contributions included updating datasets and existing tools to work with said datasets, as well as aiding in the design and planning of experiments, and manuscript writing. The model that resulted from this research was able to achieve up to 2x inference speed and ~2-4% higher mAP compared to RetinaNet with an EfficientNet backbone on the task of PCB component detection. This research was awarded Best Vision Paper at the 8<sup>th</sup> Annual Conference on Vision and Intelligent Systems.
- Research supervised by Dr. Alexander Wong.

## Deep Learning Research Assistant

January 2022 - April 2022

*Bader Lab, University of Toronto*

- Using language models, developed NLP tools for identifying PubMed articles with biomolecular interactions of potential interest. This enabled the contacting of authors whose articles would be suitable for Biofactoid at scale. Accelerating the development of this web service, which is a crucial long-term project for the Bader Lab.
- Preprocessed training and test data for these language models using confident learning and the Cleanlab library to help quantify and rectify hundreds of labelling errors.
- Supervised by Dr. Gary Bader.

## Data Science Research Assistant

May 2021 - August 2021

*Lunenfeld-Tanenbaum Research Institute, Temerty Faculty of Medicine, University of Toronto*

- Using R, identified important leukocyte biomarkers to focus on when researching improved methods of testing for and treating granulomatosis with polyangiitis (GPA). Calculated feature importance using ElasticNet and random forest models, and aggregated results using Borda's method.
- Ensured that ElasticNet and random forest models performed well in general and compared to other machine learning classifiers using nested cross-validation in R and Python (Scikit-Learn, Pandas).
- Presented findings to researchers and professors in various faculties including immunology and computer science, securing support for further research.
- Supervised by Dr. Katherine Siminovitch.

## Machine Learning R&D Intern

January 2020 - December 2020

*IQVIA, Montreal, Quebec*

- Developed a proof of concept for a completely automated record linkage system which employed cost-sensitive machine learning algorithms. This system could be used to significantly improve the speed at which data on healthcare professionals could be provided to clients by eliminating the need for manual intervention.
- Developed a pipeline to query an Amazon DynamoDB database and preprocess training and test data for this system.
- Improved system scalability through distributed computing using PySpark.
- Presented my work to developers and senior leadership across Canada through information sessions that I organized and hosted.

## Publications:

- **Palayew, S.**, Wang, B.\* , & Bader, G.\* (2025). *Towards Applying Large Language Models to Complement Single-Cell Foundation Models*. In arXiv [cs.LG]. arXiv. <http://arxiv.org/abs/2507.10039>.
- Xie, R., **Palayew, S.**, Toma, A., Bader, G., & Wang, B. (2024). *WangLab at MEDIQA-M3G 2024: Multimodal Medical Answer Generation using Large Language Models*. In T. Naumann, A. Ben Abacha, S. Bethard, K. Roberts, & D. Bitterman (Eds.), *Proceedings of the 6th Clinical Natural Language Processing Workshop* (pp. 624–634). Association for Computational Linguistics.
- Toma, A., Xie, R., **Palayew, S.**, Lawler, P., & Wang, B. (2024). *WangLab at MEDIQA-CORR 2024: Optimized LLM-based Programs for Medical Error Detection and Correction*. In T. Naumann, A. Ben Abacha, S. Bethard, K. Roberts, & D. Bitterman (Eds.), *Proceedings of the 6th Clinical Natural Language Processing Workshop* (pp. 616–623). Association for Computational Linguistics.
- Li, B.\* , **Palayew, S.\***, Li, F., Abbasi, S., Nair, S., & Wong, A. (2022). *PCBDet: An Efficient Deep Neural Network Object Detection Architecture for Automatic PCB Component Detection on the Edge* [Presentation]. 8<sup>th</sup> Annual Conference on Vision and Intelligent Systems. Waterloo, ON, Canada. **Awarded Best Vision Paper.**

\* Equal Contribution